Periodicity Transforms

William A. Sethares and Thomas W. Staley

***Abstract*—This paper presents a method of detecting periodicities in data that exploits a series of projections onto "periodic subspaces." The algorithm finds its own set of nonorthogonal basis elements (based on the data), rather than assuming a fixed predetermined basis as in the Fourier, Gabor, and wavelet transforms. A major strength of the approach is that it is linear-in-period rather than linear-in-frequency or linear-in-scale. The algorithm is derived and analyzed, and its output is compared to that of the Fourier transform in a number of examples. One application is the finding and grouping of rhythms in a musical score, another is the separation of periodic waveforms with overlapping spectra, and a third is the finding of patterns in astronomical data. Examples demonstrate both the strengths and weaknesses of the method.**

I. Introduction

HUMANS are very good at identifying complex patterns; the auditory system easily senses intricate periodicities such as the rhythms that normally occur in music and speech, the visual system readily grasps the symmetries and repetitions inherent in textures and tilings, the mind searches for simple underlying regularities to explain phenomena that appear complex and irregular. Computers are comparatively poor at locating such patterns, although several kinds of "transforms" are aimed at automatically identifying underlying patterns. Perhaps the best known is the Fourier transform, which attempts to explain a data set as a weighted sum of sinusoidal basis elements. When the data can be closely approximated by such elements, the Fourier transform provides both an explanation and a concise representation of the data. Similarly, the wavelet transform decomposes data into a sum of basis elements that are defined by a family of scaling functions. Again, when the data has the assumed scaling similarities, this provides a convincing representation and explanation. None of the available methods, however, directly searches for periodicities, repetitions, or regularities in the data. This paper builds on a classical technique called the Buys–Ballot table for the determination of "hidden periodicities" [23] to fill this gap. By analogy with modern wavelet and other transforms, we call this the *periodicity transform* (PT).

The PT decomposes sequences into a sum of periodic sequences by projecting onto a set of "periodic subspaces" $\mathcal{P}_p$, leaving residuals whose periodicities have been removed. As the name suggests, this decomposition is accomplished directly in terms of periodic sequences and not in terms of frequency or scale, as do the Fourier and wavelet transforms [20]. As a consequence, the representation is linear-in-period, rather than linear-in-frequency or linear-in-scale. Unlike most transforms, the set of basis vectors is not specified *a priori*, rather, the PT finds its own "best" set of basis elements. In this way, it is analogous to the approach of Karhunen–Loeve [5], which transforms a signal by projecting onto an orthogonal basis that is determined by the eigenvectors of the covariance matrix. In contrast, the periodic subspaces $\mathcal{P}_p$ lack orthogonality, which underlies much of the power of (and difficulties with) the PT. Technically, the collection of all periodic subspaces forms a *frame* [6], which is a more-than-complete spanning set. The PT specifies ways of sensibly handling the redundancy by exploiting some of the general properties of the periodic subspaces.

The next section examines the structure of the periodic subspaces. Section III shows how to project onto periodic subspaces and uses the residuals of these projections to demonstrate relationships between the subspaces. Special care must be taken with periodicities of incommensurable lengths. Section IV describes how the PT deals with uniqueness issues caused by the nonorthogonality of the periodic subspaces. Section V presents some of the various algorithmic options available when using the PT, and analogies are drawn with a number of methods designed to remove redundancy from frames. Then, Section VI gives a series of examples that demonstrate the application of the PT to a variety of problems including the rhythmic parsing of a musical score, the separation of waveforms, the finding of a harmonic template, and a search for periodicity in astronomical data. The casual reader may wish to jump ahead to Section VI, which motivates the algorithm by showing the kinds of results possible with the PT. The final section concludes that the PT may find use in a number of applications in which periodicities are believed to exist.

II. Periodic Subspaces

A sequence of real numbers $x(k)$ is called *p-periodic* if there is an integer p with $x(k+p) = x(k)$ for all integers k. Let

$\mathcal{P}_p$ be the set of all p-periodic sequences, and

$\mathcal{P}$ be the set of all periodic sequences.

In practice, a data vector x contains N elements. This can be considered to be a single period of an element $x_N \in \mathcal{P}_N \subset \mathcal{P}$, and the goal is to locate smaller periodicities within x_N, should they exist. The strategy is to "project" x_N onto the subspaces $\mathcal{P}_p$ for $p < N$. When x_N is "close to" some periodic subspace $\mathcal{P}_p$, then there is a p-periodic element $x_p \in \mathcal{P}_p$ that is close to the original x. This x_p is an ideal choice to use

Manuscript received June 2, 1998; revised March 21, 1999. The associate editor coordinating the review of this paper and approving it for publication was Dr. Xiang-Gen Xia.

W. A. Sethares is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706-1691 USA.

T. W. Staley is with the Science and Technology Studies Department, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA.

Publisher Item Identifier S 1053-587X(99)08302-6.

1053–587X/99$10.00 © 1999 IEEE

when decomposing x. To make these ideas concrete, it is necessary to understand the structure of the various spaces, and to investigate how the needed calculations can be realized.

Observe that $\mathcal{P}_p$ is closed under addition since the sum of two sequences with period p is itself p-periodic. Similarly, $\mathcal{P}$ is closed under addition since the sum of x_1 with period p_1 and x_2 with period p_2 has period (at most) p_1p_2. Thus, with scalar multiplication defined in the usual way, both $\mathcal{P}_p$ and $\mathcal{P}$ form linear vector spaces, and $\mathcal{P}$ is equal to the union of the $\mathcal{P}_p$.

For every period p and every "time shift" s, define the sequence $\delta_p^s(j)$ for all integers j by

$$\delta_p^s(j) = \begin{cases} 1, & \text{if } (j-s) \bmod p = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The sequences δ_p^s for $s = 0, 1, 2, \ldots, p-1$ are called the p-periodic basis vectors since they form a basis for $\mathcal{P}_p$.

Example 2.1: For $p = 4$, the 4-periodic basis vectors, shown at the bottom of the page, span the 4-periodic subspace $\mathcal{P}_4$.

An inner product can be imposed on the periodic subspaces by considering the function from $\mathcal{P} \times \mathcal{P}$ into $\Re$ defined by

$$\langle x, y \rangle = \lim_{k\to\infty} \frac{1}{2k+1} \sum_{i=-k}^{k} x(i)y(i) \tag{2}$$

for arbitrary elements x and y in $\mathcal{P}$. For the purposes of calculation, observe that if $x \in \mathcal{P}_{p_1}$ and $y \in \mathcal{P}_{p_2}$, then the product sequence $x(i)y(i) \in \mathcal{P}_{p_1p_2}$ is p_1p_2-periodic, and (2) is equal to the average over a single period, that is

$$\langle x, y \rangle = \frac{1}{p_1p_2} \sum_{i=0}^{p_1p_2-1} x(i)y(i). \tag{3}$$

To see that (2) actually defines an inner product [16] on $\mathcal{P}$, there are four conditions that must be verified:

Commutativity: $\langle x, y \rangle = \langle y, x \rangle$

Additivity: $\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

Scalar Multiplication: $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$

Positivity: $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ iff $x = 0$.

The first and third conditions are obvious from (2). To demonstrate additivity

$$\begin{aligned} \langle x+y, z \rangle &= \lim_{k\to\infty} \frac{1}{2k+1} \sum_{i=-k}^{k} (x(i)+y(i))z(i) \\ &= \lim_{k\to\infty} \frac{1}{2k+1} \sum_{i=-k}^{k} x(i)z(i) \\ &\quad + \lim_{k\to\infty} \frac{1}{2k+1} \sum_{i=-k}^{k} y(i)z(i) \\ &= \langle x, z \rangle + \langle y, z \rangle. \end{aligned}$$

The first part of the positivity condition is clear since $\langle x, x \rangle$ is a sum of squares. Suppose that $x \neq 0$ has period p. Then, $x^2(j) = \epsilon > 0$ for at least one j in each period. Then

$$\langle x, x \rangle = \lim_{k\to\infty} \frac{1}{2k+1} \sum_{i=-k}^{k} x^2(i) \geq \frac{\epsilon}{p} > 0$$

and the only periodic x with $\langle x, x \rangle = 0$ is $x = 0$.

Thus, (2) defines an inner product on $\mathcal{P}$, and the induced norm is

$$\|x\| = \sqrt{\langle x, x \rangle}. \tag{4}$$

The norm (4) is well suited to the investigation of periodic sequences since it gives the same value whether x is considered to be an element of $\mathcal{P}_p$, of $\mathcal{P}_{kp}$ (for positive integers k), or of $\mathcal{P}$.

Example 2.2: Let $x \in \mathcal{P}_3$ be the 3-periodic sequence $\{\ldots, 1, 2, 3, \ldots\}$, and let $y \in \mathcal{P}_6$ be the 6-periodic sequence $\{\ldots, 1, 2, 3, 1, 2, 3, \ldots\}$. Using (4), $\|x\| = \|y\|$.

As usual, the signals x and y in $\mathcal{P}$ are said to be orthogonal if $\langle x, y \rangle = 0$.

Example 2.3: The periodic basis elements δ_p^s for $s = 0, 1, \ldots, p-1$ are orthogonal, and $\|\delta_p^s\| = \sqrt{1/p}$.

The idea of orthogonality can also be applied to subspaces. A signal x is orthogonal to the subspace $\mathcal{P}_p$ if $\langle x, x_p \rangle = 0$ for all $x_p \in \mathcal{P}_p$, and two subspaces are orthogonal if every vector in one is orthogonal to every vector in the other. Unfortunately, the periodic subspaces $\mathcal{P}_p$ are not orthogonal to each other.

Example 2.4: If p_1 and p_2 are mutually prime, then

$$\langle \delta_{p_1}^s, \delta_{p_2}^s \rangle = \langle \delta_{p_1p_2}^s, \delta_{p_1p_2}^s \rangle = \frac{1}{p_1p_2} \neq 0.$$

Suppose that $p_1p_2 = p_3$. Then, $\mathcal{P}_{p_1} \subset \mathcal{P}_{p_3}$, and $\mathcal{P}_{p_2} \subset \mathcal{P}_{p_3}$, which restates the fact that any sequence that is p-periodic is also np-periodic for any integer n. However, $\mathcal{P}_{p_3}$ can be strictly larger than $\mathcal{P}_{p_1} \cup \mathcal{P}_{p_2}$.

Example 2.5: Let $x = \{\ldots, 1, 2, 1, -1, -2, -1, \ldots\} \in \mathcal{P}_6$. Then, x is orthogonal to both $\mathcal{P}_2$ and $\mathcal{P}_3$ since direct calculation shows that x is orthogonal to δ_2^s and to δ_3^s for all s.

In fact, no two subspaces $\mathcal{P}_p$ are linearly independent since $\mathcal{P}_1 \subset \mathcal{P}_p$ for every p. This is because the vector $\mathbf{1}$ (the 1-periodic vector of all ones) can be expressed as the sum of the p periodic basis vectors

$$\mathbf{1} = \sum_{s=0}^{p-1} \delta_p^s$$

for every p. In fact, $\mathcal{P}_1$ is the only commonality between $\mathcal{P}_{p_1}$ and $\mathcal{P}_{p_2}$ when p_1 and p_2 are mutually prime. More generally, we have the following theorem.

j	$\cdots$	-4	-3	-2	-1	0	1	2	3	4	5	6	7	$\cdots$
$\delta_4^0(j)$	$\cdots$	1	0	0	0	1	0	0	0	1	0	0	0	$\cdots$
$\delta_4^1(j)$	$\cdots$	0	1	0	0	0	1	0	0	0	1	0	0	$\cdots$
$\delta_4^2(j)$	$\cdots$	0	0	1	0	0	0	1	0	0	0	1	0	$\cdots$
$\delta_4^3(j)$	$\cdots$	0	0	0	1	0	0	0	1	0	0	0	1	$\cdots$

Theorem 2.1: $\mathcal{P}_{np} \cap \mathcal{P}_{mp} = \mathcal{P}_p$ when n and m are mutually prime.

A proof is in the Appendix. Theorem 2.1 shows how the structure of the periodic subspaces reflects the structure of the integers.

III. Projection onto Periodic Subspaces

The primary reason for formulating this problem in an inner product space is to exploit the projection theorem. Let $x \in \mathcal{P}$ be arbitrary. Then, a minimizing vector in $\mathcal{P}_p$ is an $x_p^* \in \mathcal{P}_p$ such that

$$\|x - x_p^*\| \leq \|x - x_p\|, \quad \text{for all } x_p \in \mathcal{P}_p.$$

Thus, x_p^* is the p-periodic vector "closest to" the original x. The projection theorem, which is stated here in slightly modified form, shows how x_p^* can be characterized as an orthogonal projection of x onto $\mathcal{P}_p$.

Theorem 3.1 (The Projection Theorem) [Luenberger]: Let $x \in \mathcal{P}$ be arbitrary. A necessary and sufficient condition that x_p^* be a minimizing vector in $\mathcal{P}_p$ is that the error $x - x_p^*$ be orthogonal to $\mathcal{P}_p$.

Since $\mathcal{P}_p$ is a finite (p-dimensional) subspace, x_p^* will in fact exist, and the projection theorem provides, after some simplification, a simple way to calculate it. The optimal $x_p^* \in \mathcal{P}_p$ can be expressed as a linear combination of the periodic basis elements δ_p^s as

$$x_p^* = \alpha_0 \delta_p^0 + \alpha_1 \delta_p^1 + \cdots + \alpha_{p-1} \delta_p^{p-1}.$$

According to the projection theorem, the unique minimizing vector is the orthogonal projection of x on $\mathcal{P}_p$, that is, $x - x_p^*$ is orthogonal to each of the δ_p^s for $s = 0, 1, \ldots, p-1$. Thus

$$0 = \langle x - x_p^*, \delta_p^s \rangle = \langle x - \alpha_0 \delta_p^0 - \alpha_1 \delta_p^1 - \cdots - \alpha_{p-1} \delta_p^{p-1}, \delta_p^s \rangle.$$

Since the δ_p^s are orthogonal to each other, this can be rewritten using the additivity of the inner product as

$$\begin{aligned} &= \langle x - \alpha_s \delta_p^s, \delta_p^s \rangle \\ &= \langle x, \delta_p^s \rangle - \alpha_s \langle \delta_p^s, \delta_p^s \rangle \\ &= \langle x, \delta_p^s \rangle - \frac{\alpha_s}{p}. \end{aligned}$$

Hence, α_s can be written as

$$\alpha_s = p \langle x, \delta_p^s \rangle.$$

Since $x \in \mathcal{P}$, it is periodic with some period N. From (3), the above inner product can be calculated

$$\alpha_s = p \frac{1}{pN} \sum_{i=0}^{pN-1} x(i) \delta_p^s(i).$$

However, δ_p^s is zero except when $(s - i) \bmod p = 0$, and threrefore, this simplifies to

$$\alpha_s = \frac{1}{N} \sum_{n=0}^{N-1} x(s + np). \tag{5}$$

If, in addition, N/p is an integer, then this reduces to

$$\alpha_s = \frac{1}{N/p} \sum_{n=0}^{N/p-1} x(s + np). \tag{6}$$

Example 3.1 $N = 14, p = 2$*:* Let

$$\begin{aligned} x = \{&\ldots, 2, -1.1, -1.1, 2, -1.2, -1.2, 2, -1.1 \\ &- 1.1, 2, -1.2, -1.1, 2, -1.1, \ldots\} \in \mathcal{P}_{14}. \end{aligned}$$

Then, the projection of x onto $\mathcal{P}_2$ is $x_2 = \{\ldots, 0.2, -0.228, \ldots\}$.

This sequence x_2 is the 2-periodic sequence that best "fits" this 14-periodic x. However, looking at this x closely suggests that it has more of the character of a 3-periodic sequence, albeit somewhat truncated in the final "repeat" of the $2, -1, -1$. Accordingly, it is reasonable to project x onto $\mathcal{P}_3$.

Example 3.2 $N = 14, p = 3$*:* Let $x \in \mathcal{P}_{14}$ be as defined in Example 3.1. Then, the projection of x onto $\mathcal{P}_3$ [using (5)] is $x_3 = -0.2\{\ldots, 1, 1, 1, \ldots\}$.

Clearly, this is not in accord with the intuition that this x is "almost" 3-periodic. In fact, this is an example of a rather generic effect. Whenever N and p are mutually prime, the sum in (5) cycles through all the elements of x, and therefore, $\alpha_s = \frac{1}{N} \sum_{i=0}^{N-1} x(i)$ for all s. Hence, the projection onto $\mathcal{P}_p$ is the vector of all ones (times the mean value of the x). The problem here is the incommensurability of the N and p.

What does it mean to say that x (with length N) is p-periodic when N/p is not an integer? Intuitively, it should mean that there are $\lfloor N/p \rfloor$ complete repeats of the p-periodic sequence (where $\lfloor z \rfloor$ is the largest integer less than or equal to z) plus a "partial repeat" within the remaining $\bar{N} = N - p\lfloor N/p \rfloor$ elements. For instance, the $N = 14$ sequence

$$x_1, x_2, x_3, x_1, x_2, x_3, x_1, x_2, x_3, x_1, x_2, x_3, x_1, x_2$$

can be considered a (truncated) 3-periodic sequence.

There are two ways to formalize this notion: to "shorten" x so that it is compatible with p or to "lengthen" δ_p^s so that it is compatible with N. Although it is roughly equivalent (they differ only in the final $\bar{N}$ elements), the first approach is simpler since it is possible to replace x with $x_{\tilde{N}}$ (the $\tilde{N}$-periodic sequence constructed from the first $\tilde{N} = p\lfloor N/p \rfloor$ elements of x) whenever the projection operator is involved. With this understanding, (5) becomes

$$\alpha_s = \frac{1}{\lfloor N/p \rfloor} \sum_{n=0}^{\lfloor N/p \rfloor - 1} x_{\tilde{N}}(s + np). \tag{7}$$

Example 3.3 $N = 14, p = 3$*:* Let $x \in \mathcal{P}_{14}$ be as defined in Example 3.1. Then, the projection of x onto $\mathcal{P}_3$ [using (7)] is $x_3 = \{\ldots, 2, -1.14, -1.125, \ldots\}$.

Clearly, this captures the intuitive notion of periodicity far better than Example 3.2, and the sum (7) forms the foundation of the PT. The calculation of each α_s thus requires $\lfloor N/p \rfloor$ operations (additions). Since there are p different values of s, the calculation of the complete projection x_p requires $\tilde{N} \approx N$ additions. A subroutine that carries out the needed calculations is available at our web site [24].

Let $\pi(x, \mathcal{P}_p)$ represent the projection of x onto $\mathcal{P}_p$. Then

$$\pi(x, \mathcal{P}_p) = \sum_{s=0}^{p-1} \alpha_s \delta_p^s \tag{8}$$

where the δ_p^s are the (orthogonal) p-periodic basis elements of $\mathcal{P}_p$. Clearly, when $x \in \mathcal{P}_p$, $x = \pi(x, \mathcal{P}_p)$. By construction, when x is projected onto $\mathcal{P}_{np}$, it finds the best np-periodic components within x, and hence, the residual $r = x - \mathcal{P}_{np}$ has no np-periodic component. The content of the next result is that this residual also has no p-periodic component. In essence, the projection onto $\mathcal{P}_{np}$ "grabs" all the p-periodic information.

Theorem 3.2: For any integer n, let $r = x - \pi(x, \mathcal{P}_{np})$ be the residual after projecting x onto $\mathcal{P}_{np}$. Then, $\pi(r, \mathcal{P}_p) = 0$.

All proofs are found in the Appendix. The next result relates the residual after projecting onto $\mathcal{P}_p$ to the residual after projection onto $\mathcal{P}_{np}$.

Theorem 3.3: Let $r_p = x - \pi(x, \mathcal{P}_p)$ be the residual after projecting x onto $\mathcal{P}_p$. Similarly, let $r_{np} = x - \pi(x, \mathcal{P}_{np})$ denote the residual after projecting x onto $\mathcal{P}_{np}$. Then

$$r_{np} = r_p - \pi(r_p, \mathcal{P}_{np}).$$

Combining the two previous results shows that the order of projections does not matter in some special cases, that is

$$\pi(x, \mathcal{P}_p) = \pi(\pi(x, \mathcal{P}_p), \mathcal{P}_{np}) = \pi(\pi(x, \mathcal{P}_{np}), \mathcal{P}_p)$$

which is used in Section V to help sensibly order the projections.

IV. Nonuniqueness

The previous section shows how to project an arbitrary signal $x \in \mathcal{P}$ onto $\mathcal{P}_p$, that is, how to calculate $x_p = \pi(x, \mathcal{P}_p)$. It is not completely obvious, however, how to iterate this calculation to reliably decompose x into its constituent periodic elements. There are several subtleties, most of which are related to the lack of orthogonality of the subspaces $\mathcal{P}_p$.

Most standard transforms can be interpreted as projections onto suitable subspaces, and in most cases (such as the Fourier and wavelet transforms), the subspaces are orthogonal. Such orthogonality implies that the projection onto one subspace is independent of the projection onto others. Thus, a projection onto one sinusoidal basis function (in the Fourier transform) is independent of the projections onto others, and the Fourier decomposition can proceed by projecting onto one subspace, subtracting out the projection, and repeating. Orthogonality guarantees that the order of projection is irrelevant. This is not true for projection onto nonorthogonal subspaces such as the periodic subspaces $\mathcal{P}_p$.

Example 4.1: Let $x = \{\ldots, 1, 0, 4, -3, 4, 0, \ldots\} \in \mathcal{P}_6$.

a) If x is projected onto $\mathcal{P}_6$ to give b_6, then the projections of $x - b_6$ onto $\mathcal{P}_2$ and $\mathcal{P}_3$ are zero since $x = b_6$.
b) If x is projected onto $\mathcal{P}_2$ to give $b_2 = \{\ldots, 3, -1, \ldots\}$, then the projection of $x - b_2$ onto $\mathcal{P}_3$ is $b_3 = \{\ldots, -2, 1, 1, \ldots\}$. The resulting decomposition is $x = b_2 + b_3$.
c) If x is projected onto $\mathcal{P}_3$ to give $\bar{b}_3 = \{\ldots, -1, 2, 2, \ldots\}$, then the projection of $x - \bar{b}_3$ onto $\mathcal{P}_2$ is $\bar{b}_2 = \{\ldots, 2, -2, \ldots\}$. The resulting decomposition is $x = \bar{b}_2 + \bar{b}_3$.

Thus, the order in which the projections occur affects the decomposition, and the PT does not in general provide a unique representation. Once the succession of the projections is specified, then the answer is unique, and the next section details a handful of ways to specify how the projections can be meaningfully ordered. For instance, they can be ordered sequentially from large to small or from small to large by how much power they remove from the signal, by how much "area" they remove, or by the magnitude of their projection on independent basis elements.

This nonuniqueness may appear problematical. On the contrary, we claim that it is a strength of the PT method; that lack of uniqueness is mainly a theoretical and not a practical issue. The goal of signal decomposition is to represent the original signal in an alternative and more meaningful way. The Fourier transform provides one representation, the Gabor transform another, and wavelet transforms give a different decomposition for each "mother wavelet" chosen. The PT can provide several different decompositions into periodic basis elements, depending on how the projections are ordered. If the method of ordering is well suited to the task at hand, then the resulting decomposition can convincingly reflect the underlying meaning of the signal. At a deeper level, many real-world phenomena are fundamentally ambiguous. For instance, the way humans perceive rhythm in music [11] or "connect the dots" in a visual scene can be essentially ambiguous [10], and the various criteria of merit used to order the projections in the PT may lead to multiple solutions that reflect this essential ambiguity.

V. Algorithms for Periodic Decomposition

The PT searches for the best periodic characterization of the length N signal x. The underlying technique is to project x onto some periodic subspace giving $x_p = \pi(x, \mathcal{P}_p)$, which is the closest p-periodic vector to x. This periodicity is then removed from x, leaving the residual $r_p = x - x_p$ stripped of its p-periodicities. Both the projection x_p and the residual r_p may contain other periodicities and may be decomposed into other q-periodic components by projection onto $\mathcal{P}_q$. The trick in designing a useful algorithm is to provide a sensible criterion for choosing the order in which the successive p's and q's are chosen. The intended goal of the decomposition, the amount of computational resources available, and the measure of "goodness-of-fit" all influence the algorithm. The analysis of the previous sections can be used to guide the decomposition by exploiting the relationship between the structure of the various $\mathcal{P}_p$. For instance, it makes no sense to project x_p onto $\mathcal{P}_{np}$ because $x_p \in \mathcal{P}_{np}$, and no new information is obtained. This section presents several different algorithms, discusses their properties, and then compares these algorithms with some methods available in the literature.

One subtlety in the search for periodicities is related to the question of appropriate boundary (end) conditions. Given the signal x of length N, it is not particularly meaningful to look for periodicities longer than $p = N/2$, even though nothing in

the mathematics forbids it. Indeed, a "periodic" signal with length $N-1$ has $N-1$ degrees of freedom and surely can match x very closely, yet provides neither a convincing explanation nor a compact representation of x. Consequently, we restrict further attention to periods smaller than $N/2$.

Probably the simplest useful algorithm operates from small periods to large:

Small To Large Algorithm

pick threshold $T \in (0,1)$
let $r = x$
for $p = 2, 3, \ldots, N/2$
  $x_p = \pi(r, \mathcal{P}_p)$
  if $\frac{\|r - x_p\|}{\|x\|} > T$
    $r = r - x_p$
    save x_p as basis element
  end
end

The Small To Large algorithm is simple because there is no need to further decompose the basis elements x_p; if there were significant q-periodicities within x_p (where "significant" is determined by the threshold T), then they would already have been removed by x_q at an earlier iteration. The algorithm works well because it tends to favor small periodicities, to concentrate the power in $\mathcal{P}_p$ for small p, and, hence, to provide a compact representation.

Thinking of the norm as a measure of power, the threshold is used to insure that each chosen basis element removes at least a factor T of the power from the signal. Of course, choosing different thresholds leads to different decompositions. If T is chosen too small (say zero), then the decomposition will simply pick the first linear independent set from among the p-periodic basis vectors

$$\overbrace{\delta_2^1, \delta_2^2}^{\mathcal{P}_2}, \overbrace{\delta_3^1, \delta_3^2, \delta_3^3}^{\mathcal{P}_3}, \overbrace{\delta_4^1, \delta_4^2, \delta_4^3, \delta_4^4}^{\mathcal{P}_4}, \delta_5^1, \delta_5^2, \ldots$$

which defeats the purpose of searching for periodicities. If T is chosen too large, then too few basis elements may be chosen (none as $T \to 1$). In between "too small" and "too large" is where the algorithm provides interesting descriptions. For many problems, $0.01 < T < 0.1$ is appropriate, since this allows detection of periodicities containing only a few percent of the power yet ignores those p's that only incidentally contribute to x.

An equally simple "Large To Small" algorithm is not feasible because projections onto x_p for composite p may mask periodicities of the factors of p. For instance, if $x_{100} = \pi(x, \mathcal{P}_{100})$ removes a large fraction of the power, this may in fact be due to a periodicity at $p = 20$, yet further projection of the residual onto $\mathcal{P}_{20}$ is futile since $\pi(x - x_{100}, \mathcal{P}_{20}) = 0$ by Theorem 3.2. Thus, an algorithm that decomposes from large p to smaller p must further decompose both the candidate basis element x_p as well as the residual r_p since either might contain smaller q-periodicities.

The M-Best Algorithm deals with these issues by maintaining lists of the M best periodicities and the corresponding basis elements. The first step is to build the initial list as shown at the bottom of the page.

At this stage, the algorithm has compiled a list of the M periodicities q_i that remove the most "energy" (in the sense of the norm measure) from the sequence. Typically, however, the q_i will be large (since, by Theorem 3.2, the projections onto larger subspaces np contain the projections onto smaller subspaces p). Thus, the projections x_{q_i} can be further decomposed into their constituent periodic elements to see if these smaller (sub)periodicities removes more energy from the signal than another currently on the list. If so, then the new one replaces the old. It is not necessary to search all possible periods $p < q_i$ when decomposing, however; we need only search the factors. Let $\rho_i = \{n; q_i/n \text{ is an integer}\}$ be the set of factors of q_i. Then, the algorithm proceeds as shown at the bottom of the next page.

This second step projects each x_{q_i} onto each of its factors $Q \in \rho_i$. If the norm of the new projection x_Q is larger than the smallest norm in the list, and if the sum of all the norms will increase by replacing x_{q_i}, then the new Q is added to the list, and the last element x_{q_M} is deleted. These steps rely heavily on Theorem 3.3. For example, suppose that the algorithm has found a strong periodicity in (say) $\mathcal{P}_{140}$, giving the projection $x_{140} = \pi(x, \mathcal{P}_{140})$. Since $140 = 2^2 \cdot 5 \cdot 7$, the factors are $\rho = \{2, 4, 5, 7, 10, 14, 20, 28, 35, 70\}$. Then, the inner loop in step 2 searches over each of the $\pi(x_{140}, \mathcal{P}_Q)\ \forall Q \in \rho$. If x_{140} is "really" composed of a significant periodicity at (say) 20, then this new periodicity is inserted in the list and will later be searched for yet smaller periodicities. The M-Best Algorithm is relatively complex, but it removes the need for a threshold parameter by maintaining the list. This is a sensible approach and it often succeeds in building a good decomposition of the signal. A variation called the M-Best algorithm with γ-modification (or M-Best$_\gamma$) is described in Appendix B, where the measure of energy removed is normalized by the (square root of) the length p.

Another approach is to project x onto all the periodic basis elements δ_p^s for all p and s, essentially measuring the correlation between x and the individual periodic basis elements. The p with the largest (in absolute value) correlation is then used for the projection. This idea leads to the Best

---

M-Best Algorithm (step 1)

pick size M
let $r_0 = x$
for $i = 1, 2, \ldots, M$
  find q_i with $\|\pi(r_{i-1}, \mathcal{P}_{q_i})\| \geq \|\pi(r_{i-1}, \mathcal{P}_q)\|\ \forall q \in [1, 2, \ldots, N/2]$
  $r_i = r_{i-1} - \pi(r_{i-1}, \mathcal{P}_{q_i})$
  concatenate q_i and $x_{q_{i-1}} = \pi(r_i, \mathcal{P}_{q_i})$ onto respective lists
end

Correlation Algorithm.

Best Correlation Algorithm

$M =$ number of desired basis elements
let $r = x$
for $i = 1, 2, \ldots, M$
  $\rho = \operatorname{argmax}_p |\langle r, \delta_p^s \rangle|$
  save $x_\rho = \pi(r, \mathcal{P}_\rho)$ as basis element
  $r = r - x_\rho$
end

The presumption behind the "Best Correlation" algorithm is that good p will tend to have good correlation with at least one of the p-periodic basis vectors. This method tends to pick out periodicities with large regular spikes over those that are more uniform. A fourth approach is to determine the best periodicity p by Fourier methods and then to project onto $\mathcal{P}_p$, shown at the bottom of the next page.

Using frequency to find periodicity is certainly not always the best idea, but it can work well and has the advantage that it is a well-understood process. The interaction between the frequency and periodicity domains can be a powerful tool, especially since the Fourier methods have good resolution at high frequencies (small periodicities) while the PT has better resolution at large periodicities (low frequencies).

As far as we know, there is no simple way to guarantee that an optimal decomposition has been obtained. One foolproof method for finding the best M subspaces would be to search all of the possible $\binom{N}{M}$ different orderings of projections to find the one with the smallest residual. This is computationally prohibitive in all but the simplest settings, although an interesting special case is when $M = 1$, that is, when only the largest periodicity is of importance.

Several methods for finding the "best" basis functions from among some (possibly large) set of potential basis elements have been explored in the literature [6], many of which are related to variants of general "projection pursuit" algorithms [12]. Usually, these are set in the context of choosing a representation for a given signal from among a family of prespecified frame elements. For instance, a Fourier basis, a collection of Gabor functions, a wavelet basis, and a wavelet packet basis may form the elements of an over-complete "dictionary." Coifman [8] proposes an algorithm to chose a particular basis to represent a given signal based on a measure of entropy. In [17], a greedy algorithm called "matching-pursuit" is presented that successively decomposes a signal by picking the element that best correlates with the signal, subtracts off the residual, and decomposes again. This is analogous to (though somewhat more elaborate than) the "Best Correlation" algorithm above. Nafie [18] proposes an approach that maintains "active" and "inactive" dictionaries. Elements are swapped into the active dictionary when they better represent the signal than those currently active. This is analogous to the "M-Best" algorithm above. The "best basis" approach of [13] uses a thresholding method aimed at signal enhancement and is somewhat analogous to the Small To Large algorithm above. Using an l^1 norm, Chen and Donoho [7] propose a method that exploits Karmarkar's interior point linear programming method. The "method of frames" [9] essentially calculates the pseudo-inverse of a (large rectangular) matrix composed of all the vectors in the dictionary.

While these provide analogous approaches to the problems of dealing with a redundant spanning set, there are two distinguishing features of the PT. The first is that the p-periodic basis elements are inherently coupled together. For instance, it does not make any particular sense to chose (say) δ_3^1, δ_4^3, δ_7^3, and δ_9^2 as a basis for the representation of a periodic signal. The p-periodic basis elements are fundamentally coupled together, and none of the methods were designed to deal with such a coupling. More generally, none of the methods is able (at least directly) to exploit the kind of structure (for instance, the containment of certain subspaces and the equality of certain residuals) that is inherent when dealing with the periodic subspaces of the PT.

VI. Examples

This section gives three examples of the application of the PT to the automatic grouping of rhythmic motifs in a musical score, to the separation of periodic signals when given their sum, and to the identification of periodicities in certain astronomical data.

A. Musical Rhythm

One situation where periodicities play a key role is in musical rhythm. Several researchers (for instance, [2], [15] and

---

M-Best Algorithm (step 2)

repeat until no change in list
  for $i = 1, 2, \ldots, M$
    find Q^* with $\|\pi(x_{q_i}, \mathcal{P}_{Q^*})\| \geq \|\pi(x_{q_i}, \mathcal{P}_Q)\|\ \forall Q \in \rho_i$
    let $x_{Q^*} = \pi(x_{q_i}, \mathcal{P}_{Q^*})$ be the projection onto $\mathcal{P}_{Q^*}$
    let $x_{q^*} = x_{q_i} - x_{Q^*}$ be the residual
    if $\|x_{q^*}\| + \|x_{Q^*}\| > \|x_{q_M}\| + \|x_{q_i}\|$ & $\|x_{q^*}\| > \min_k \|x_{q_k}\|$ & $\|x_{Q^*}\| > \min_k \|x_{q_k}\|$
      replace q_i with q^* and x_{q_i} with x_{q^*}
      insert Q^* and x_{Q^*} into lists at position $i - 1$
      remove q_M and x_{q_M} from end of lists
    end if
  end for
end repeat

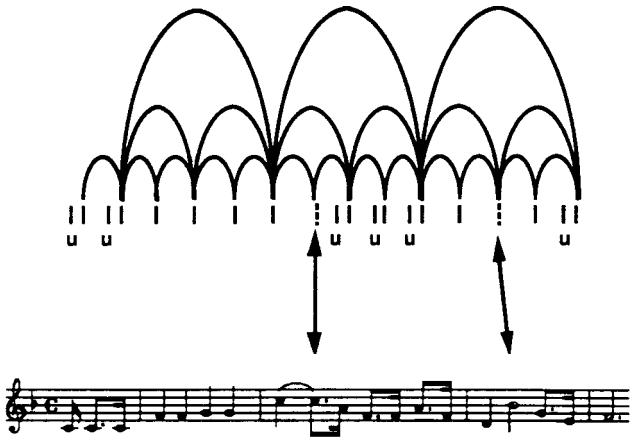

Fig. 1. Rhythmic grouping calculated by Rosenthal's program Fa. Essentially the same grouping was achieved by the PT, using a simple coding of the onset times.

[19]) have created programs to automatically parse a musical score (or a standard MIDI file, which contains equivalent information) in order to generate a "high-level" explanation of the grouping of musical rhythms. The "pulse" is the basic unit of temporal structure and is often represented in the score as a quarter note. Such pulses are typically gathered together into groupings that correspond to metered measures, and these groupings are often clustered to form larger structures corresponding to musical "phrases." Such patterns of grouping and clustering can continue through many hierarchical levels, and many of these may be readily perceptible to attentive listeners. It is not so easy to teach computers to recognize these patterns.

Rosenthal [19] has created a rhythm parsing program called "Fa," which searches for regularly spaced onset times in a MIDI data stream. The program forms and ranks several hypotheses "according to criteria that correspond to ways in which human listeners choose rhythmic interpretations." These criteria include quite sophisticated ideas such as having accented notes fall on "strong" beats, noticing motivic repetitions, and measuring salience. An example is given of the best rhythmic parsing found by Fa for the song "La Marseillaise," which is reproduced here as Fig. 1.

The four measures of "La Marseillaise" were coded into the binary sequence

$$1\,100\,110\,001\,000\,100\,010\,001\,000\,000\,110$$
$$011\,001\,100\,010\,000\,000\,100\,110\,000\,000\,000$$

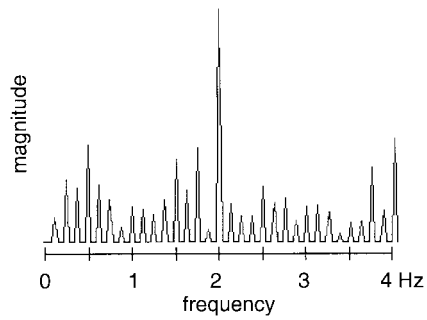

Fig. 2. DFT of the binary sequence derived from the score to La Marseillaise, assuming a sampling rate of 8 Hz (one half note per second). The peak at 2 Hz represents the quarter note pulse, and the peak at 0.5 Hz corresponds to the measure.

in which each digit represents a time equal to that of one sixteenth note. A 1 indicates that a note event occurred at that time, whereas a 0 means that no new note event occurred. This was made "periodic" by concatenating it to itself four times and then decomposed by the various periodicity algorithms.

The Small To Large algorithm with threshold 0.1 detected four periodicities: with periods 4 (corresponding to the quarter notes), 16 (the measured bar lines), 32 (the two measure phrases), and 64 (everything else). These periodicities removed 28%, 13%, 20%, and 39% of the power, respectively, and agree well with Rosenthal's analysis. Similarly, the Best Correlation algorithm (with ratio 0.1) found the three periodicities 4, 16, and 32. When asked to find four periodicities, the M-Best algorithm responded with 8, 16, 32, and 64, whereas the M-Best$_\gamma$ algorithm returned 2, 4, 16, and 32. It is typical that the γ-modified version emphasizes smaller periodicities. When asked to find the ten best periodicities (the default), both M-best versions detected 2, 4, 8, 16, 32, 64, 0, 0, 0, and 0. The final four zeroes indicate that the residual was zero and the decomposition exact. The Best Frequency algorithm returned periods 4, 16, and 5. The unexpected 5 is due to "rounding off" of the spectrum near period 4. Since the frequency corresponding to period 5 has the largest magnitude in this residual, the algorithm can proceed no further.

For comparison, the same binary sequence (again concatenated four times) was tranformed using the DFT. Assuming a sampling rate of 8 Hz, the duration of each quarter note is 0.5 s, and the resulting magnitude spectrum is shown in Fig. 2. In this figure, the largest peak at 2 Hz represents the quarter note pulse, whereas the peak at 0.5 Hz corresponds to the measure. It is unclear how to interpret the remainder of the information in this spectrum.

Best Frequency Algorithm

$M =$ number of desired basis elements
let $r = x$
for $i = 1, 2, \ldots, M$
  $y = \|DFT\{r\}\|$
  $p = \text{Round}(1/f)$, where $f =$ frequency at which y is max
  save $x_p = \pi(r, \mathcal{P}_p)$ as basis element
  $r = r - x_p$
end

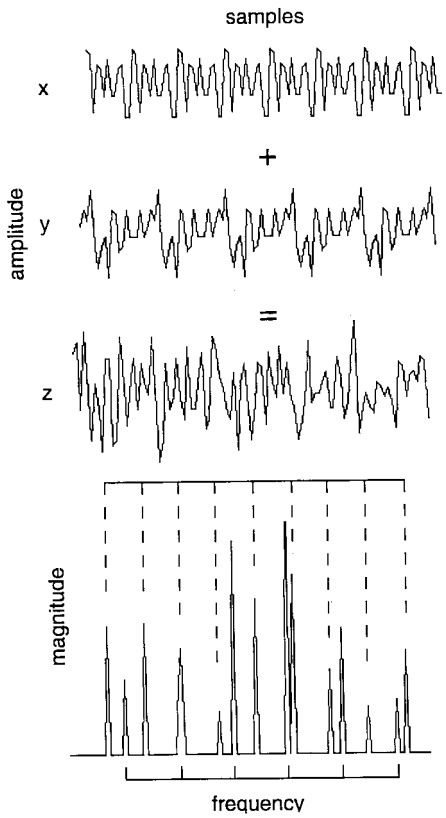

Fig. 3. Signal z is the sum of the 13-periodic x and the 19-periodic y. The spectrum shows the overlapping of the two spectra, which cannot be separated by linear filtering. The PT is able to locate the periodicities (which were *a priori* unknown) and to resconstruct (up to a constant offset) both x and y given only z.

B. Signal Separation

When signals are added together, information is often lost. However, if there is some characteristic that distinguishes the signals, then they may be recoverable from their sum. Perhaps the best known example is when the spectrum of x and the spectrum of y do not overlap. Then, both signals can be recovered from $x+y$ with a linear filter. However, if the spectra overlap significantly, the situation is more complicated. This example shows how, if the underlying signals are periodic in nature, then the PT can be used to recover signals from their sum. This process can be thought of as a way to extract a "harmonic template" from a complicated spectrum.

Consider the signal z in Fig. 3, which is the sum of the 13-periodic signal x and the 19-periodic signal y. The spectrum of z is quite complex, and it is not obvious which parts of the spectrum arise from x and which from y. However, when the PT is applied to z, two periodicities are found, at 13 and at 19, with basis elements that are exactly $x_{13} = x + c_1$ and $y_{19} = y + c_2$, that is, both signals x and y are recovered, up to a constant. Thus, the PT is able to locate the periodicities (which were assumed *a priori* unknown) and to reconstruct (up to a constant offset) both x and y given only their sum. Even when z is contaminated with 50% random noise, the PT still locates the two periodicities, although the reconstructions of x and y are noisy. To see the mechanism, let η be the noise signal, and let $\eta_{13} = \pi(\eta, \mathcal{P}_{13})$ be the projection of η onto the 13-periodic subspace. The algorithm then finds $x_{13} = x + c_1 + \eta_{13}$ as its 13-periodic basis element.

The four transforms do not behave identically. The Small To Large algorithm regularly finds such periodic sequences. The Best Correlation algorithm works best when the periodic data is spiky. The M-Best algorithm is sometimes fooled into returning multiples of the basic periodicities (say 26 or 39 instead of 13), whereas the M-Best$_\gamma$ is overall the most reliable and noise resistant. The Best Frequency algorithm often becomes "stuck" when the frequency with the largest magnitude does not closely correspond to an integer periodicity. The behaviors of the algorithms are explored in detail in the four demonstration files (*PTdemos2l*, *PTdemobc*, *PTdemomb*, and *PTdemobf*) that accompany the periodicity software [24].

Two aspects of this example deserve comment. First, the determination of the periodicity and its corresponding basis element is tantamount to locating a "harmonic template" in the frequency domain. For example, the 13-periodic component has a spectrum consisting of a fundamental (at a frequency f_1 proportional to $1/13$) and harmonics at $2f_1, 3f_1, 4f_1, \ldots$. Similarly, the 19-periodic component has a spectrum consisting of a fundamental (at a frequency f_2 proportional to $1/19$) and harmonics at $2f_2, 3f_2, 4f_2, \ldots$. These are indicated in Fig. 3 by the beams above and below the spectrum of z. Thus, the PT provides a way of finding simple harmonic templates that may be obscured by the inherent complexity of the spectrum, and the process of subtracting the projection from the original signal can be interpreted as a multinotched filter that removes the relevant fundamental and its harmonics. For a single p, this is a kind of "gapped weight" filter familiar from time series analysis [14].

The offsets c_1 and c_2 occur because $\mathcal{P}_1$ is contained in both $\mathcal{P}_{13}$ and in $\mathcal{P}_{19}$. In essence, both of these subspaces are capable of removing the constant offset (which is an element of $\mathcal{P}_1$) from z. Were both x and y zero mean, then both c_1 and c_2 would be zero, but since both are nonzero mean, the projection onto (say) $\mathcal{P}_{13}$ grabs all of the signal in $\mathcal{P}_1$ for itself (Thus, $c_1 = \text{mean}(x) + \text{mean}(y)$, and further projection onto $\mathcal{P}_{19}$ gives $c_2 = -\text{mean}(y)$). This illustrates a general property of projections onto periodic subspaces. Suppose that the periodic signals to be separated were $x_{np} \in \mathcal{P}_{np}$ and $x_{mp} \in \mathcal{P}_{mp}$ for some mutually prime n and m. Since $\mathcal{P}_{np} \cap \mathcal{P}_{mp}$ is $\mathcal{P}_p$, both $\mathcal{P}_{np}$ and $\mathcal{P}_{mp}$ are capable of representing the common part of the signal, and x_{np} and x_{mp} can only be recovered up to their common component in $\mathcal{P}_p$. In terms of the harmonic templates, there is overlap between the set of harmonics of x_{np} and the harmonics of x_{mp}, and the algorithm does not know whether to assign the overlapping harmonics to x_{np} or to x_{mp}.

It is also possible to separate a deterministic periodic sequence $z \in \mathcal{P}_p$ from a random sequence y when only their sum $x = y + z$ can be observed. Suppose that y is a stationary (independent, identically distributed) process with mean m_y. Then, $E\{\pi(y, \mathcal{P}_p)\} = m_y \cdot \mathbf{1}$ (where $\mathbf{1}$ is the vector of all ones), and therefore

$$E\{\pi(x, \mathcal{P}_p)\} = E\{\pi(y+z, \mathcal{P}_p)\} = E\{\pi(y, \mathcal{P}_p)\} + E\{\pi(z, \mathcal{P}_p)\} = m_y \cdot \mathbf{1} + z$$

since $E\{\pi(z, \mathcal{P}_p)\} = E\{z\} = z$. Hence, the deterministic periodicity z can be identified (up to a constant) and removed from x. Such decomposition will likely be most valuable when there is a strong periodic "explanation" for z and, hence, for x. In some situations such as economic and geophysical data sets, regular daily, monthly, or yearly cycles may obscure the underlying signal of interest. Projecting onto the subspaces $\mathcal{P}_p$, where p corresponds to these known periodicities is very sensible. However, appropriate values for p need not be known *a priori*. By searching through an appropriate range of p (exploiting the various algorithms of Section V), both the value of p and the best p-periodic basis element can be recovered from the data itself.

C. Patterns in Astronomical Data

To examine the performance of the PT in the detection of more complex patterns, a 3-min segment of astronomical data gathered by the Voyager spacecraft (published on audio CD in [21]) was analyzed. When listening to this CD, there is an apparent pulse rate with approximately 18 (not necessarily equal length) pulses in each 32-s segment. Because of the length of the data, significant downsampling was required. This was accomplished by filtering the digital audio data in overlapping sections and calculating the rms value in each section. The resulting sequence approximates the amplitude of the Voyager signal with an effective sampling rate of $\frac{44100}{2520} = 17.5$ Hz.

The downsampled data was first analyzed with a Fourier transform. The most significant sinusoidal components occur at 0.078, 0.137, 0.157, 0.167, 0.177, and 0.216 Hz, which correspond to periodicities at 12.8, 7.3, 6.3, 6.0, 5.7 and 4.6 s. Because the Fourier transform is linear-in-frequency, the values are less accurate at long periods (low frequencies). For example, whereas the time interval between adjacent Fourier bins is only ± 0.2 s for the shortest of the significant periodicities (4.6 s), the time between bins at the longest detected periodicity (12.8 s) is approximately ± 1.6 s.

Applying the PT to the downsampled data using the M-Best$_\gamma$ algorithm (with $M = 10$) gives the data at the bottom of the page. The shortest periodicity at 1.77 s corresponds well with the pulse rate that is apparent when listening to the data directly (the large norm for this first subspace results from a significant DC term). The structure of the results mimics the operation of the algorithm. For example, there are three sets of related periodicities. The first three 31:217:434 are in the ratio 1:7:14, the next two 124:868 are in the ratio 1:7 (and are also in the ratio 4:28 relative to the periodicity 31). The third set 328:656 are in the ratio 1:2. These are indicative of the decomposition of large periods by Step 2 of the M-Best$_\gamma$ algorithm. While these longer periodicities are not immediately obvious when "listening to" or "looking at" at the data, their importance is reinforced by a comparison with the Fourier results. The PT periodicities at 7.09 and 12.4 correspond to the Fourier periodicities at 7.3 and 12.8 (recall the inherent margin of error in the bin width of the FFT). In addition, the periodicity at 12.4 contains the 2:1 subperiod at 6.2 detected by the FFT. The 1.77-s pulse is approximately one third of the 5.7-s Fourier result, reinforcing the interpretation that this was an underlying "fundamental."

On the other hand, the appearance of several periodicities without mutual factors clustered at one time scale (i.e., the periodicities at 45.66, 50.29 and 49.6 s), suggests that one long periodicity in the data may have been inexactly decomposed into several related components. This relates to what may be the most severe limitation to the general applicability of the PT: When the sample interval does not correspond to a factor of some periodicity in the data, the decomposition of small periods from larger ones is difficult. Qualitatively, this can be thought of as the converse of the limitation of the Fourier method; while the linear-in-frequency behavior of the FFT increases the error at long periods, the linear-in-period behavior of the PT causes inefficiencies at short periods. Just as an increase in the amount of data used in the FFT can provide better precision, a decrease in the sample interval can improve the performance of the PT. Nonetheless, there is no general method of ensuring *a priori* that a particular periodicity in unknown data will correspond to a multiple of the sample rate. To see this effect clearly, we resampled the data at an effective sampling interval of 0.065 s and then reapplied the M-Best$_\gamma$ algorithm. In this case, the longer periods were not as successfully decomposed. Similar sensitivities to the choice of sampling rates were observed in an early critique of the Buys–Ballot method [4].

This "integer periodicity" limitation of the PT can be mitigated by the proper choice of a highly factorable integer as the sample interval. In general, the identification of small periodicities within an arbitrary data set will be most efficient when the sample interval itself contains many factors (many exact periodicities). These intervals, each composed of many periodic "building blocks," are easily combined to identify larger multiples. In fact, this was the reason we chose the effective sampling rate based on a subsampling interval of 2520, which factors as $2^3 \cdot 3^2 \cdot 5 \cdot 7$. This interval has a highest density of factors within the desired range of effective sampling rates (0.05 s to 0.1 s) consistent with a downsampled data set of reasonable length.

VII. Conclusion

The PT is designed to locate periodicities within a data set by projecting onto the (nonorthogonal) periodic subspaces. The method decomposes signals into their basic periodic

Period p	31	217	434	124	868	328	656	799	880	525
Time (seconds)	1.77	12.4	24.8	7.09	49.6	18.74	37.49	45.66	50.29	30
Norm (in percent)	29.6	6.2	4.3	2.8	8.6	3.7	5.9	4.9	3.6	2.7

components, creating its own "basis elements" as linear combinations of delta-like p-periodic basis vectors.

In some cases, the PT can provide a clearer explanation of the underlying nature of the signals than standard techniques. For instance, the signal z of Fig. 3 is decomposed into (roughly) 14 complex sinusoids by the DFT or into two periodic sequences by the PT. In a strict mathematical sense, they are equivalent since the residuals are equal in norm. However, the PT "explanation" is simpler and allows the recovery of the individual elements from their sum. When periodicity provides a better explanation of a signal or an event than does frequency, then the PT is likely to outperform the DFT. Conversely, when the signal incorporates clear frequency relationships, the DFT will likely provide a clearer result. Our belief is that a detailed analysis of truly unknown signals will benefit from the application of all available techniques. Several examples of possible applications of the PT are given:

1) to the grouping of rhythmic patterns in a musical score;
2) to the separation of signals;
3) to the finding of harmonic templates;
4) to the search for patterns in astronomical data.

There are several ways that these ideas can be extended. Generalizations to two dimensions are straightforward, and there is some hope that visual patterns such as textures may be amenable to identification and classification with the PT. The algorithms and results were all stated in terms of real numbers, but the generalization to the complex setting is certainly feasible. For simplicity, the development was conducted with discrete sequences rather than functions, but the ideas may be extendible to function spaces using "almost periodic functions" [3]. Another intriguing idea is to incorporate certain kinds of *a priori* knowledge about the signal within the decomposition procedure, for instance, if projections are allowed only onto the periodic subspaces

$$\mathcal{P}_2 \subset \mathcal{P}_4 \subset \mathcal{P}_8 \subset \mathcal{P}_{16} \subset \ldots$$

then there is a nesting of subspaces with orthogonal complements as is familiar from the wavelet theory [20].

Like the Hadamard transform [22], the PT can be calculated using only additions (no multiplications are required). As shown in Section III, each projection requires approximately N operations, but the calculations required to project onto (say) $\mathcal{P}_p$ overlap the calculations required to project onto $\mathcal{P}_{np}$ in a nontrivial way, and these redundancies can undoubtedly be exploited in a more efficient implementation.

The analysis of this paper focused on the basic properties of the periodic subspaces $\mathcal{P}_p$ and of projections onto $\mathcal{P}_p$. There may be better ways to parse the periodic subspaces than those presented in Section V, and there is room for exploring the convergence properties of the algorithms, their robustness to noises, and to the interactions between periodicities, frequencies, and sampling rates.

Appendix A

Theorem 2.1: $\mathcal{P}_{np} \cap \mathcal{P}_{mp} = \mathcal{P}_p$ when n and m are mutually prime.

Proof: Observe first that if $x \in \mathcal{P}_p$, then it is also in both $\mathcal{P}_{np}$ and $\mathcal{P}_{mp}$ since a p-periodic sequence is also kp-periodic for any k. On the other hand, suppose that $x \in \mathcal{P}_{np} \cap \mathcal{P}_{mp}$. Then, x is both np-periodic and mp-periodic, that is,

$$x(i) = x(i+np) = x(i+mp) \quad \forall i.$$

This means that $x(i) = x(i+(jn+km)p)$ for all integers j and k. In particular, since m and n are mutually prime, the function kn mod m cycles through all possible values $[0, 1, 2, \ldots, m-1]$ as k cycles through the first m integers. Hence, there is a k^* and a j^* such that $j^*n + k^*m = 1$. Hence, $x(i) = x(i+(j^*n+k^*m)p) = x(i+p)$, and x is p-periodic. □

The next two results make use of the projection operator $\pi(x, \mathcal{P}_p)$ defined in (8). Because of the additivity of the inner product

$$\pi(ax+by, \mathcal{P}_p) = a\pi(x, \mathcal{P}_p) + b\pi(y, \mathcal{P}_p)$$

for any $x, y \in \mathcal{P}$ and any real numbers $a, b \in \Re$. The parameters $\alpha_{s_{\tilde{N}}}$ of the projections are calculated as in (7), although the $\tilde{N}$ notation is suppressed.

Theorem 3.2: For any integer n, let $r = x - \pi(x, \mathcal{P}_{np})$ be the residual after projecting x onto $\mathcal{P}_{np}$. Then, $\pi(r, \mathcal{P}_p) = 0$.

Proof:

$$\begin{aligned}
\pi(r, \mathcal{P}_p) &= \pi(x - \pi(x, \mathcal{P}_{np}), \mathcal{P}_p) \\
&= \pi(x, \mathcal{P}_p) - \pi(\pi(x, \mathcal{P}_{np}), \mathcal{P}_p) \\
&= \pi(x, \mathcal{P}_p) - \pi\left(np \sum_{i=0}^{np-1} \langle x, \delta_{np}^i \rangle \delta_{np}^i, \mathcal{P}_p\right) \\
&= \pi(x, \mathcal{P}_p) - p \sum_{s=0}^{p-1} \left\langle np \sum_{i=0}^{np-1} \langle x, \delta_{np}^i \rangle \delta_{np}^i, \delta_p^s \right\rangle \delta_p^s.
\end{aligned}$$

Since

$$\langle \delta_{np}^i, \delta_p^s \rangle = \begin{Bmatrix} 1/np, & \text{if } i \bmod p = s \\ 0, & \text{otherwise} \end{Bmatrix}$$

all the "cross terms" are annihilated, and this simplifies to

$$= \pi(x, \mathcal{P}_p) - p \sum_{s=0}^{p-1} \left\langle \sum_{m=0}^{n-1} \langle x, \delta_{np}^{s+mp} \rangle \delta_{np}^{s+mp}, \delta_p^s \right\rangle \delta_p^s.$$

Observe that

$$\begin{aligned}
\sum_{m=0}^{n-1} \langle a_{s+mp} \delta_{np}^{s+mp}, \delta_p^s \rangle &= \sum_{m=0}^{n-1} a_{s+mp} \langle \delta_{np}^{s+mp}, \delta_p^s \rangle \\
&= \frac{1}{np} \sum_{m=0}^{n-1} a_{s+mp}.
\end{aligned}$$

Equating $a_{s+mp} = \langle x, \delta_{np}^{s+mp} \rangle$, and substituting gives

$$= \pi(x, \mathcal{P}_p) - p \sum_{s=0}^{p-1} \sum_{m=0}^{n-1} \langle x, \delta_{np}^{s+mp} \rangle \delta_p^s$$
$$= \pi(x, \mathcal{P}_p) - p \sum_{s=0}^{p-1} \left\langle x, \sum_{m=0}^{n-1} \delta_{np}^{s+mp} \right\rangle \delta_p^s.$$

The np-periodic basis vectors "sum up" to a p-periodic basis vector, that is

$$\sum_{m=0}^{n-1} \delta_{np}^{s+mp} = \delta_p^s$$

and the expression finally simplifies to

$$= \pi(x, \mathcal{P}_p) - p \sum_{s=0}^{p-1} \langle x, \delta_p^s \rangle \delta_p^s = \pi(x, \mathcal{P}_p) - \pi(x, \mathcal{P}_p) = 0.$$

□

Theorem 3.3: Let $r_p = x - \pi(x, \mathcal{P}_p)$ be the residual after projecting x onto $\mathcal{P}_p$. Similarly, let $r_{np} = x - \pi(x, \mathcal{P}_{np})$ denote the residual after projecting x onto $\mathcal{P}_{np}$. Then

$$r_{np} = r_p - \pi(r_p, \mathcal{P}_{np}).$$

Proof: The above equation can be rewritten

$$x - \pi(x, \mathcal{P}_{np}) = x - \pi(x, \mathcal{P}_p) - \pi(r_p, \mathcal{P}_{np}).$$

Substituting into the third projection and expanding gives

$$x - \pi(x, \mathcal{P}_{np}) = x - \pi(x, \mathcal{P}_p) - \pi(x, \mathcal{P}_{np}) + \pi(\pi(x, \mathcal{P}_p), \mathcal{P}_{np}).$$

Canceling terms leaves

$$\pi(\pi(x, \mathcal{P}_p), \mathcal{P}_{np}) = \pi(x, \mathcal{P}_p)$$

which is true because the projection onto $\mathcal{P}_p$ leaves a p-periodic sequence, which is also an np-periodic sequence by definition. □

Appendix B

Suppose that $x(i)$ is a random sequence with each term distributed $N(0, \sigma)$ and independent (if the mean is not zero, then project onto $\mathcal{P}_1$ and remove the DC bias). The measure of "goodness" in the M-Best algorithm can be easily modified so that on average, no periodicity is reported any stronger than any other, as would be expected for "random" data. Thus, the measure $||\pi(x, \mathcal{P}_p)||$ in the algorithm may be replaced by $||\pi(x, \mathcal{P}_p)||/\sqrt{p}$, as shown below.

Recall from (8) and (7) that $\pi(x, \mathcal{P}_p) = \sum_{s=0}^{p-1} \alpha_s \delta_p^s$, where each α_s is a sum of $\lfloor N/p \rfloor$ of the $x(i)$'s and is therefore, distributed $N(0, \frac{\sigma}{\sqrt{\lfloor N/p \rfloor}})$. Thus, $\frac{\alpha_s}{\sigma/\sqrt{\lfloor N/p \rfloor}}$ is distributed as $N(0, 1)$, and the α_s are independent. A standard result from probability says that if $y = \xi_1^2 + \xi_2^2 + \cdots + \xi_p^2$, where each ξ_i^2 are independent $N(0, 1)$, then $\sqrt{y}$ has $\mathcal{X}$ distribution with p degrees of freedom, and the mean of $\sqrt{y}$ is $\gamma(p) = \frac{\sqrt{2}\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})}$, where Γ denotes the standard gamma function. Hence, the mean of

$$\sqrt{\sum_{i=1}^{p} \left(\frac{\alpha_i}{\sigma/\sqrt{\lfloor N/p \rfloor}} \right)^2} \approx \frac{\sqrt{N}}{\sigma\sqrt{p}} \sqrt{\sum_{i=1}^{p} \alpha_i^2} = \frac{\sqrt{N}}{\sigma} ||\pi(x, \mathcal{P}_p)||$$

is also $\gamma(p)$. Thus, in the algorithm, rather than comparing $||\pi(x, \mathcal{P}_p)||$ with $||\pi(x, \mathcal{P}_q)||$ directly, a better comparison is between $\frac{||\pi(x, \mathcal{P}_p)||}{\gamma(p)}$ and $\frac{||\pi(x, \mathcal{P}_q)||}{\gamma(q)}$. Since [1]

$$\Gamma(az + b) \approx \sqrt{2\pi} e^{-az} (az)^{az+b-\frac{1}{2}}$$

$\gamma(p)$ can be approximated as $\frac{\sqrt{2}\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \approx \sqrt{p}$, which is considerably simpler numerically. Thus, we suggest using $\frac{||\cdot||}{\sqrt{p}}$ in place of $||\cdot||$ throughout the algorithm, which we call the M-Best algorithm with γ modification, or M-Best$_\gamma$ for short.

Acknowledgment

The authors would like to thank I. Dobson for periodic conversations about rhythmic matters. J. Gubner and D. Cobb helped keep the authors from becoming Lost in (inner product) Space.

References

[1] Abramowitz and Stegun, *Handbook of Mathematical Fuctions*. New York: Dover, 1972.
[2] J. Brown, "Determination of the meter of musical scores by autocorrelation," *J. Acoust. Soc. Amer.*, vol. 94, no. 4, pp. 1953–1957, Oct. 1993.
[3] H. Bohr, *Almost Periodic Functions*. New York: Chelsea, 1947.
[4] H. Burkhardt, *Trigonometrische Interpolation*, Enzykl. Math. Wiss., Bd II, Teil 1, 1904.
[5] J. B. Burl, "Estimating the basis functions of the Karhunen-Loeve transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 99–105, Jan. 1989.
[6] C. S. Burrus, R. A. Gopinath, and H. Guo, *Wavelets and Wavelet Transforms*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[7] S. Chen and D. L. Donoho, "Basis pursuit," in *Proc. 28th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1994, pp. 41–44.
[8] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best-basis selection," *IEEE Trans. Inform. Theory*, vol. 38, Mar. 1992.
[9] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Inform. Theory*, vol. 34, pp. 605–612, July 1988.
[10] J. J. Gibson, *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin, 1966.
[11] S. Handel, "Using polyrhythms to study rhythm," *Music Percep.*, vol. 1, no. 4, pp. 465–484, Summer 1984.
[12] P. J. Huber, "Projection pursuit," *Ann. Stat.*, vol. 13, no. 2, pp. 435–475, 1985.
[13] H. Krim, S. Mallat, D. Donoho, and A. Willsky, "Best basis algorithm for signal enhancement," in *Proc. IEEE Conf. Acoust., Speech Signal Process.*, Detroit, MI, May 1995, pp. 1561–1564.
[14] L. H. Koopmans, *The Spectral Analysis of Time Series*. San Diego, CA: Academic, 1995.
[15] H. C. Longuet-Higgins and C. S. Lee, "The rhythmic interpretation of monophonic music," *Music Percep.*, vol. 1, no. 4, pp. 424–441, Summer 1984.
[16] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1968.
[17] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, Dec. 1993.

[18] M. Nafie, M. Ali, and A. Tewfik, "Optimal subset selection for adaptive signal representation," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, Atlanta, GA, May 1996, pp. 2511–2514.
[19] D. Rosenthal, "Emulation of rhythm perception," *Comput. Music J.*, vol. 16, no. 1, Spring 1992.
[20] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley, MA: Wellesley, 1996.
[21] *Symphonies of the Planets: NASA Voyager Recordings*, Los Angeles CA: Delta, 1993.
[22] R. K. R. Yarlagadda and J. E. Hershey, *Hadamard Matrix Analysis and Synthesis*. Boston, MA: Kluwer, 1997.
[23] H. Wold, *Stationary Time Series*. Uppsala, Sweden: Almquist and Wiskells, 1938, republished 1954.
[24] MATLAB programs for the calculation of the Periodicity Transforms [Online] Available http://eceserv0.ece.wisc.edu/~sethares.

William A. Sethares received the B.A. degree in mathematics from Brandeis University, Waltham, MA, and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY.

He has worked at the Raytheon Company as a Systems Engineer and is currently Associate Professor with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison. His research interests include adaptation and learning in signal processing, communications, and acoustics, and he is the author of *Tuning, Timbre, Spectrum, Scale* (New York: Springer, 1997).

Thomas W. Staley received the B.S. degree in chemical engineering from the University of Delaware, Newark, in 1989, and the M.S. and Ph.D. degrees in materials science from the University of Wisconsin, Madison, in 1991 and 1997, respectively.

He is currently pursuing additional graduate work with the Science and Technology Studies Department, Virginia Polytechnic Institute and State University, Blacksburg. His research interests include the investigation of structural order in materials, computer modeling, and interpretation of physical phenomena, the history and philosophy of perception and related technological issues, and the physics of musical instruments.